Data Center / Data Contact

- Author(s):
Kathleen Carr, Erich Reiter

Data Center / Data Contact

- Author(s): Kathleen Carr, Erich Reiter

Description of Problem

- Roles:
    - pick lists include roles that apply to organizations and roles that apply to individuals
    - role values are not the same for DIF and for ECHO
    - current collection records in CMR have non-standardized values for roles (including different variations of the same word (ARCHIVER, archive, archive data center), different words for same concept (PRODUCER, PROCESSOR?)
    - ECHO 10 records have Organization role built into field names (Processing Center, Archive Center) rather than as field values
    - EDSC shows Processing Center and Archiving Center on Collection Information panel, but these are blank for DIF collections
- Contact Types:
    - contact type lists associated with organizations and personnel are not standardized
    - contact type lists associated with organizations and personnel do not include 'modern' values
- Normalization:
    - organization and personnel information are repeated in multiple CMR collection records; in GCMD, they are normalized

    (The normalization issues will be addressed later, in

⚠ ECSE-91 - JIRA project doesn't exist or you don't have permission to view it.

and

⚠ ECSE-99 - JIRA project doesn't exist or you don't have permission to view it.

)

JIRA Linkage

⚠ *ECSE-92 - JIRA project doesn't exist or you don't have permission to view it.*

⚠ *ECSE-75 - JIRA project doesn't exist or you don't have permission to view it.*

Background

Recommendations

1. **Roles**

   a. **RECOMMENDATION 1**: Distinguish between Roles for Organizations (call them Data Centers) and Roles for Personnel or Personnel Groups (call them Data Contacts) , i.e.,
   Change ResponsibilityRoleEnum to:

   DataCenterRoleEnum and

   DataContactRoleEnum

   Current list in UMM-Common schema ResponsibilityRoleEnum mixes the three:

   "enum": ["RESOURCEPROVIDER", "CUSTODIAN", "OWNER", "USER", "DISTRIBUTOR", "ORIGINATOR", "POINTOFCONTACT", "PRINCIPALINVESTIGATOR", "PROCESSOR", "PUBLISHER", "AUTHOR", "SPONSOR", "COAUTHOR", "COLLABORATOR", "EDITOR", "MEDIATOR", "RIGHTSHOLDER", "CONTRIBUTOR", "FUNDER", "STAKEHOLDER"] (NOTE: This corresponds to the CI_RoleCode code list in the UMM-Common document (Figure 3).

   See also http://www.ngdc.noaa.gov/metadata/published/xsd/schema/resources/Codelist/gmxCodelists.xml#CI_RoleCode

   b. **RECOMMENDATION 2**: Assign the VALUES to DataCenterRoleEnum and DataContactRoleEnum as follows:

   i. For **Organizations** (Data Centers), use: Archiver , Processor, Distributor, and Originator

   i.e., DataCenterRoleEnum = "enum": ["ARCHIVER", "DISTRIBUTOR", "ORIGINATOR", "PROCESSOR"]

   (Note: GCMD Organization Type ENUM is:

   < xs:enumeration value="DISTRIBUTOR"/>

   < xs:enumeration value="ARCHIVER"/>

   < xs:enumeration value="ORIGINATOR"/>

   < xs:enumeration value="PROCESSOR"/>

   ⚠ CMR-2706 - JIRA project doesn't exist or you don't have permission to view it. Error because Organization was set to ARCHIVER.

   ii. for **Personnel (DataContact)**, use:

   DataContactRoleEnum = "enum": ["DATA CENTER CONTACT", "TECHNICAL CONTACT", "SCIENCE CONTACT", "INVESTIGATOR","METADATA AUTHOR", "USER SERVICES", "SCIENCE SOFTWARE"]

NOTE: The **GCMD Role ENUMs** are at https://git.earthdata.nasa.gov/projects/EMFD/repos/dif-schemas/browse/10.x/UmmCommon.xsd

The **GCMD Personnel Role ENUM** is:

<xs:enumeration value="INVESTIGATOR"/>

< xs:enumeration value="INVESTIGATOR, TECHNICAL CONTACT"/>

 <xs:enumeration value="METADATA AUTHOR"/>

< xs:enumeration value="METADATA AUTHOR, TECHNICAL CONTACT"/>

< xs:enumeration value="TECHNICAL CONTACT"/>

NOTE: The **GCMD OrganizationPersonnelRole ENUM** is   < xs:enumeration value="DATA CENTER CONTACT"/>

⚠ ECSE-98 - JIRA project doesn't exist or you don't have permission to view it.

   c. **RECOMMENDATION 3**: Map current role **values** in the CMR to the new proposed role values as follows:

Current distribution of Role Values

Organization Role mapping:

Current values:(may need to check with provider in some cases to confirm whether mapping is satisfactory).

Count of Role Row Labels	Total
DIF10	1462
ARCHIVER	913
DISTRIBUTOR	183
ORIGINATOR	182
PROCESSOR	184
DIF9	31258
ARCHIVER	31258

ECHO10	15743
Archive	889
ARCHIVER	599
archiving data center	138
author	71
CUSTODIAN	4526
DATA CENTER CONTACT	2005
Data Manager	271
Data Originator	128
DIF AUTHOR	738
DIF AUTHOR, TECHNICAL CONTACT	8
DISTRIBUTOR	3
GHRC USER SERVICES	349
internal data center	138
INVESTIGATOR	936
INVESTIGATOR, TECHNICAL CONTACT	1
PROCESSOR	3254
Producer	613
TECHNICAL CONTACT	577
TECHNICAL CONTACT, DIF AUTHOR	28
User Services	471

Proposed value mapping:

UMM-C	DIF 9	DIF 10	ECHO 10	ISO
ARCHIVER	ARCHIVER	ARCHIVER	ARCHIVER Archive archiving data center CUSTODIAN Data Manager internal data center	
DISTRIBUTOR		DISTRIBUTOR	DISTRIBUTOR	
ORIGINATOR		ORIGINATOR	Data Originator author	
PROCESSOR		PROCESSOR	PROCESSOR Producer	

Personnel Role mapping:

Current values:

Count of Role	
Row Labels	Total
⊟ EOP	867
Associate Principal Investigator	4
Deputy Science Software Development Manage	42
GLAS Science Team Leader	42
ICESat Project Scientist	42
NSIDC USER SERVICES	3
Primary Contact	3
Product Team Leader	3
Science Software Development Manager.	42
Sea Ice Algorithms	3
Snow Algorithms	3
TECHNICAL CONTACT	560
Technical Contact for Science	60
User Services	60
⊟ OCG	545
DATA CENTER CONTACT	545
⊟ OCP	917
DATA CENTER CONTACT	917
⊟ OP	31882
DATA CENTER CONTACT	31817
INVESTIGATOR	65
⊟ P	53199
DATA CENTER CONTACT	38
DIF AUTHOR	19327
INVESTIGATOR	17803
METADATA AUTHOR	6
TECHNICAL CONTACT	16025
⊟ PCG	2
TECHNICAL CONTACT	2
⊟ PCP	2763
INVESTIGATOR	725
METADATA AUTHOR	985
TECHNICAL CONTACT	1053
Grand Total	###

Proposed value mapping: (may need to check with provider in some cases to confirm whether mapping is appropriate).

UMM-C	CURRENT VALUE			
"DATA CENTER CONTACT",	DATA CENTER CONTACT Primary Contact			
"TECHNICAL CONTACT",	TECHNICAL CONTACT Product Team Leader			
"SCIENCE CONTACT",	Technical Contact for Science GLAS Science Team Leader ICESAT Project Scientist			
"INVESTIGATOR",	INVESTIGATOR Associate Principal Investigator			
"METADATA AUTHOR",	DIF AUTHOR METADATA AUTHOR TECHNICAL CONTACT, DIF AUTHOR			
"USER SERVICES"	NSIDC USER Services User Services GHRC USER SERVICES			

"SCIENCE SOFTWARE"	Science Software Development Manager			
	Deputy Science Software Development Manager			
	Sea Ice Algorithms ❓			
	Snow Algorithms ❓			

⚠ ECSE-75 - JIRA project doesn't exist or you don't have permission to view it.

⚠ UMMC-412 - JIRA project doesn't exist or you don't have permission to view it.

⚠ UMMC-435 - JIRA project doesn't exist or you don't have permission to view it.

d. **RECOMMENDATION 4:**

Map ECHO 10 Organization metadata to UMM-C as follows:

/Collection/ProcessingCenter=<value> maps to DataCenter/Role=PROCESSOR and DataCenter/ShortName = <shortname corresponding to Processing Center value) and DataCenter/LongName = <longname corresponding to Processing Center value)

/Collection/ArchiveCenter= <value> maps to DataCenter/Role=ARCHIVER and DataCenter/ShortName = <shortname corresponding to Archive Center value) and DataCenter/LongName = <longname corresponding to Archive Center value)

⚠ CMR-600 - JIRA project doesn't exist or you don't have permission to view it.

ISSUE: Write a ticket to determine how to map the <value> for Processing Center or Archive Center in the ECHO record to the corresponding Shortname and Longname (NEED A LIST OF ACTUAL VALUES IN ECHO 10 RECORDS FOR PROCESSING CENTER AND ARCHIVE CENTER)

EXAMPLES:

In UAT, for Collection with Collection Shortname AQUARIOUS_L4_OISSS_IPRC_7DAY_V4, Version 1:

<ArchiveCenter>PO.DAAC</ArchiveCenter> should translate to DataCenter/Role=ARCHIVER and DataCenter/ShortName = NASA/JPL/PODAAC

<ProcessingCenter>IPRC/SOEST University of Hawaii, Manoa</ProcessingCenter> should translate to DataCenter/Role=PROCESSOR and DataCenter/ShortName = UHI/SOEST/IPRC

e. **RECOMMENDATION 5:**

Map DIF 10 metadata to UMM-C as follows:

UMM-C DataCenter/Role maps to DIF 10 Organization/Organization_Type

UMM-C DataCenter/ShortName maps to DIF 10 Organization/Organization_Name/Short_Name

EXAMPLE:

<Organization><Organization_Type>DISTRIBUTOR</Organization_Type> <Organization_Name> <Short_Name>CA/EC/MSC</Short_Name> </Organization_Name> maps to

 DataCenter/Role=DISTRIBUTOR and DataCenter/ShortName = CA/EC/MSC

2.   Contact Type

RECOMMENDATION: Create a schema ENUM list which combines the current GCMD list with three additional contact types proposed by MMT

developers: Email, Facebook, Twitter

  a. The GCMD has the following ENUM list for Contact Type.

<xs:enumeration value="Direct Line"/>

<xs:enumeration value="Primary"/>

<xs:enumeration value="Telephone"/>

<xs:enumeration value="Fax"/>

<xs:enumeration value="Mobile"/>

<xs:enumeration value="Modem"/>

<xs:enumeration value="TDD/TTY Phone"/>

<xs:enumeration value="U.S. toll free"/>

<xs:enumeration value="Other"/>

This recommendation has already been implemented on the MMT using

⚠ MMT-538 - JIRA project doesn't exist or you don't have permission to view it.

Changes to UMM-C fields and UMM-Common fields:

See the proposed class diagram for the element definitions, their types and cardinality in Lucidchart (you may need to log in to Lucidchart).

The mapping from UMM-C OLD to UMM-C Proposed is in the attached spreadsheet.

⚠ CMR-2298 - JIRA project doesn't exist or you don't have permission to view it.

⚠ ECSE-75 - JIRA project doesn't exist or you don't have permission to view it.

⚠ CMR-2265 - JIRA project doesn't exist or you don't have permission to view it.

⚠ CMR-1256 - JIRA project doesn't exist or you don't have permission to view it.

⚠ UMMC-336 - JIRA project doesn't exist or you don't have permission to view it.

Mappings to DIF, ECHO, ISO

The DIF-UMM-ECHO_Mapping.xlsx file can be found here: https://git.earthdata.nasa.gov/projects/EMFD/repos/unified-metadata-model

⚠ ECSE-45 - JIRA project doesn't exist or you don't have permission to view it.

Interoperability Considerations

During the original UMM-C review, it was decided that Organization and Personnel elements would be combined into a merged element called

Responsible Party, similar to the ISO 19115-2 field CI_ResponsibleParty. Later during the UMM-Common review it was decided to separate role with party to allow for components or xlinks and for reusability of the party element.

The MMT collection forms are organized based on the current UMM-C/UMM-Common metadata model.   During MMT alpha and beta testing, some metadata providers indicated that they were confused by the Organization, Personnel and Party elements on the MMT collection record forms. Based on that feedback, we revisited the metadata model for these field groupings, giving consideration to how the metadata model can best reflect the physical model of organizations responsible for earth science datasets and people or groups who can be contacted with questions about those datasets.   Based on this further analysis, we are now recommending a model which moves away from the Organization, Responsibility, and Party concepts, and moves back to the Data Center concept that better reflects the physical model of NASA DAACs and makes translation between GCMD collection record values and UMM-C values more straightforward and intuitive.

In addition, we are recommending changing the current UMM-C Personnel element to DataContact, where 'contact' is the operative word.   The re-naming of this element to emphasize the function of a contact person, and the corresponding new role enum assigned to DataContact, better addresses a dataset user's possible need to know how to contact someone to get answers to different types of questions about the dataset (e.g., TECHNICAL CONTACT, SCIENCE CONTACT, USER SERVICES, etc.)

The current UMM-C document recommends that Roles should be validated against the ISO 19115-1 code list as shown below.

However, these ISO roles co-mingle data center roles and people (data contact) roles, and a number of the ISO roles (e.g., funder, rightsHolder) are not relevant in the context of the Data Centers and Data Contacts in the CMR earth science data sets. We recommend that instead of using the ISO roles, we use one role enum list for data centers and a separate role enum list for data contacts, as shown in the model above on this page. Our new recommended enum list for Data Center roles includes a role of ARCHIVER, which is not in the ISO list, but reflects the role of the 'Distributed Active Archive Centers' used by NASA to manage earth science data.

Mappings from DataCenter and DataContact elements to ISO elements will be provided in the UMM-C/UMM-Common documents.

Changes to CMR

The new elements will need to be added and possibly indexed if the old elements were indexed. Several names have changed and the translation code will have to be implemented.

⚠ CMR-2706 - JIRA project doesn't exist or you don't have permission to view it.

⚠ CMR-1473 - JIRA project doesn't exist or you don't have permission to view it.

⚠ CMR-1841 - JIRA project doesn't exist or you don't have permission to view it.

⚠ MMT-538 - JIRA project doesn't exist or you don't have permission to view it.

⚠ CMR-3157 - JIRA project doesn't exist or you don't have permission to view it.

⚠ CMR-3158 - JIRA project doesn't exist or you don't have permission to view it.

⚠ CMR-3161 - JIRA project doesn't exist or you don't have permission to view it.

⚠ CMR-3208 - JIRA project doesn't exist or you don't have permission to view it.

⚠ CMR-3233 - JIRA project doesn't exist or you don't have permission to view it.

What should the MMT forms look like

⚠ MMT-489 - JIRA project doesn't exist or you don't have permission to view it.

⚠ MMT-538 - JIRA project doesn't exist or you don't have permission to view it.

⚠ ECSE-80 - JIRA project doesn't exist or you don't have permission to view it.

⚠ MMT-689 - JIRA project doesn't exist or you don't have permission to view it.

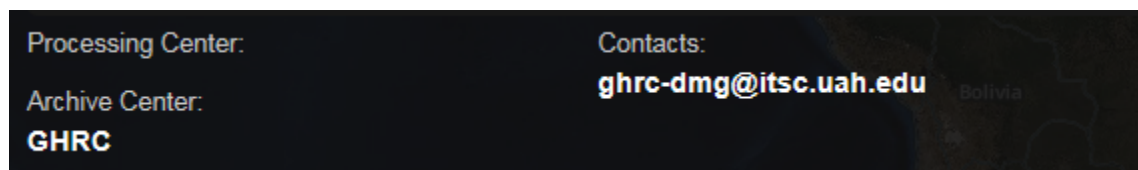⚠ MMT-690 - JIRA project doesn't exist or you don't have permission to view it.

How should the values in these fields be presented to the user on the EDSC

⚠ EDSC-999 - JIRA project doesn't exist or you don't have permission to view it.

⚠ EDSC-942 - JIRA project doesn't exist or you don't have permission to view it.

⚠ **EDSC-1141 - JIRA project doesn't exist or you don't have permission to view it.** **Organization facet** : change 'Organization' to 'Data Center' as the facet name; this reverses a decision made in ⚠ EDSC-907 - JIRA project doesn't exist or you don't have permission to view it.

On **Collection details** page, EDSC currently shows:

⚠ EDSC-1140 - JIRA project doesn't exist or you don't have permission to view it.

a. instead of 'Processing Center' and 'Archive Center', show <DataCenter/Role>:<DataCenter/ShortName> or
<DataCenter/ShortName>(<DataCenter/Role>)

e.g. ARCHIVER: GHRC or GHRC (ARCHIVER)

b. Show all Data Centers listed in the collection metadata record

c. Also display ServiceHours, Contact_Instructions, and RelatedURL for each Data Center, if these fields are present

⚠️ EDSC-1142 - JIRA project doesn't exist or you don't have permission to view it.

Show all Data Contacts, as <DataContact/Role>: DataContact/ContactPerson/First_Name> <DataContact/ContactPerson/Middle_Name> <DataContact/ContactPerson/Last_Name> |

a. <DataContact/Role>: <DataContact/ContactGroup/GroupName>

and show all contact mechanisms for the data contact

<DataContact/ContactInformation/ContactMechanism/Type>: <DataContact/ContactInformation/ContactMechanism/Value>

e. g. SCIENCE CONTACT: John Doe, Telephone: (555)1212, Email: johndoe@gmail.com

or USER SERVICES: DAAC User Services, Email: daacuserservices@daacname.gov

b. Also display ServiceHours, Contact_Instructions, and RelatedURL for each Data Contact, if these fields are present

How should pick lists / controlled vocabulary be handled?

Data Center Field	Current Source of Pick list values (MMT)	Proposed Source of Pick list values	Is this field used for EDSC Faceted Search?
DataCenter/Role	Schema ENUM - current ResponsibilityRoleEnum	Schema ENUM - proposed DataCenterRoleEnum ["ARCHIVER", "DISTRIBUTOR", "ORIGINATOR", "PROCESSOR"]	no
DataCenter/ShortName	KMS (Data Center)	KMS (Data Center)	yes
DataCenter/LongName	KMS, or auto fill from KMS after selection short name	KMS, or auto fill from KMS after selection short name	no
DataCenter/ContactInformation/ContactMechanism/Type	GCMD list plus current MMT values	Schema ENUM - proposed (GCMD list plus current MMT values)	no
DataCenter/ContactInformation/Address/Country	ISO	ISO	no
DataCenter/ContactInformation/Address/StateProvince	ISO	ISO	no
DataCenter/ContactInformation/RelatedURLs	auto fill from KMS after selecting short name or long name	auto fill from KMS after selecting short name or long name	no

Data Contact Field	Current Source of Pick list values (MMT)	Proposed Source of Pick list values	Is this field used for EDSC Faceted Search?
DataContact/Role DataCenter/DataContact/Role	Schema ENUM - current ResponsibilityRoleEnum	Schema ENUM - proposed DataContactRoleEnum	no
DataContact/ContactInformation/ContactMechanism/Type DataCenter/DataContact/ContactInformation/ContactMechanism/Type	GCMD list plus current MMT values	Schema ENUM - proposed (GCMD list plus current MMT values)	no
DataContact/ContactInformation/Address/Country DataCenter/DataContact/ContactInformation/Address/Country	ISO	ISO	no
DataContact/ContactInformation/Address/StateProvince DataCenter/DataContact/ContactInformation/Address/StateProvince	ISO	ISO	no

Reconciliation of Existing Metadata values with new rules

⚠ UMMC-427 - JIRA project doesn't exist or you don't have permission to view it.

Map values per Recommendation 3

Standardize DataCenter shortname, longname, url per the KMS values (MMT can assist the user to do this one record at a time).

Correct organization roles that are really personnel roles

Impact of Changes on Ingest of Granule Metadata

None - Organization and Personnel information can be changed in a Collection record without impacting the Collection's granule records.

Approvals

ESDIS Decision

Choices	Your Vote	Current Result: (0 Total Votes)	Comments
Approve for Implementation	☐	0 Votes, 0%	
Approve for ESO Recommendation Only	☐	0 Votes, 0%	
Needs Additional Work	☐	0 Votes, 0%	
Reject	☐	0 Votes, 0%	

Pending Voters ✉ : dshum